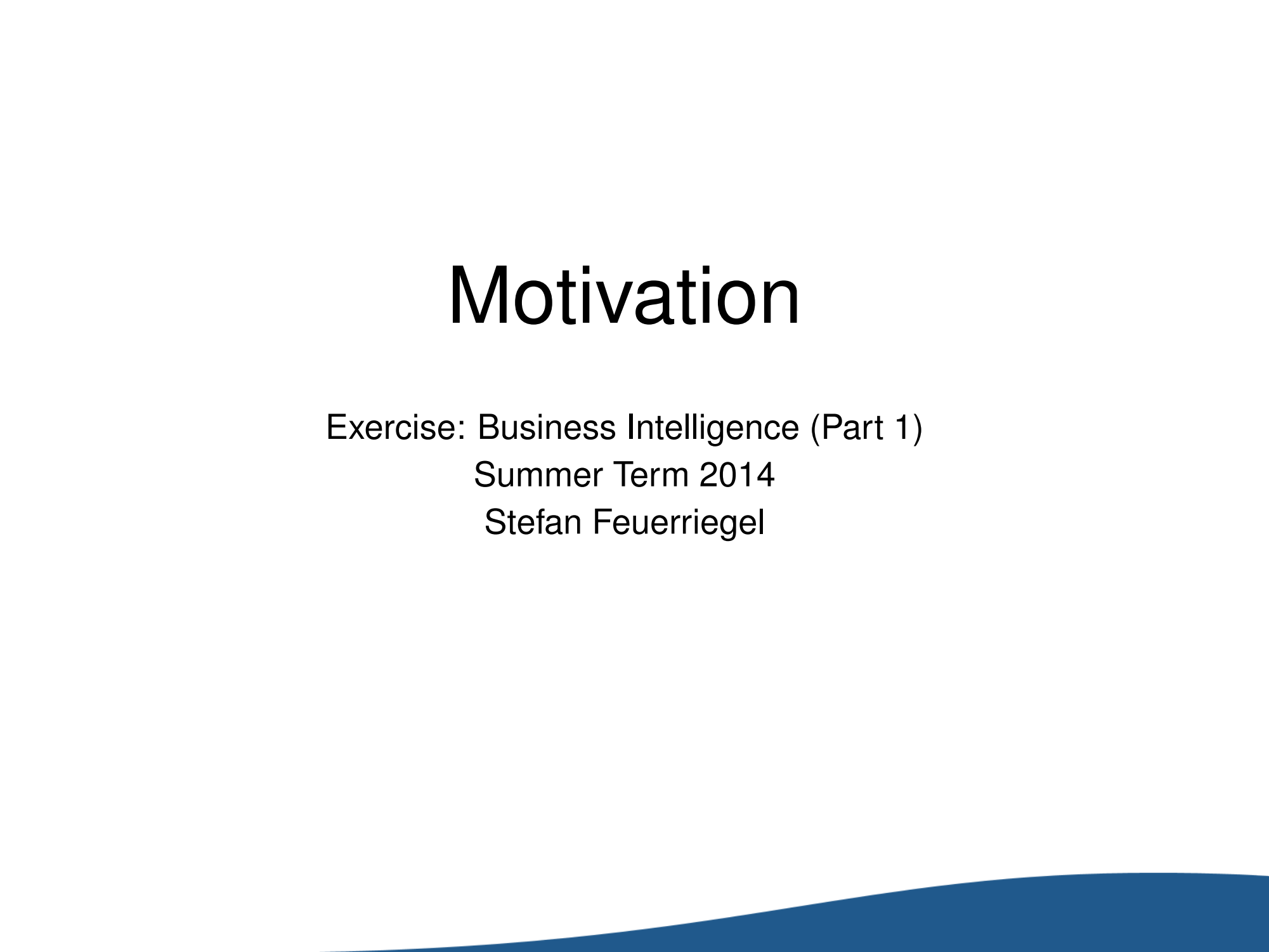
# Motivation

Exercise: Business Intelligence (Part 1)

Summer Term 2014

Stefan Feuerriegel



# Outline

# Outline

# Turning Enterprise Data into Knowledge

## Examples from Practice

- Based on historical values, how can businesses predict future developments ahead of time?
- Given the current stock market prices, can we predict tomorrow's values?
- How does weather impact electricity prices?
- Which parameters of second-hand cars correlate with their value?
- How can businesses group consumers into distinct categories according to their purchase behavior?

## (Monetary) benefits obvious $\Rightarrow$ difficult because Big Data

# Dimensions of Big Data

**Volume**

Data at scale
Terabytes to perabytes of data

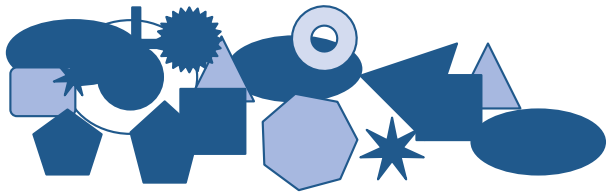
**Variety**

Data in many forms
Structured, unstructured, text, multimedia

**Velocity**

Data in motion
Analysis of streaming data to enable

**Veracity**



Data uncertainty
Managing the reliability and predictability of inherently imprecise data types

# It's All About Data

## Dimension: Volume

Enterprises are drowned in ever-growing data of all types, easily exceeding terabytes – even petabytes – of information:

- Turn 12 terabytes of tweets posted each day into improved product sentiment analysis
- Convert 350 bn annual meter readings to better predict power consumption

→ IBM (2012). Analytics: The real-world use of big data, Executive Report.

# It's All About Data

## Dimension: Velocity

Sometimes 2 minutes is too late: for time-sensitive processes, such as catching fraud, big data must be used as it streams into your enterprise in order to maximize its value:

- Inspect 5 m trade events that occur each day to identify potential fraud
- Analyze 500 m daily call records in real-time to predict customer churn faster

→ IBM (2012). Analytics: The real-world use of big data, Executive Report.

# It's All About Data

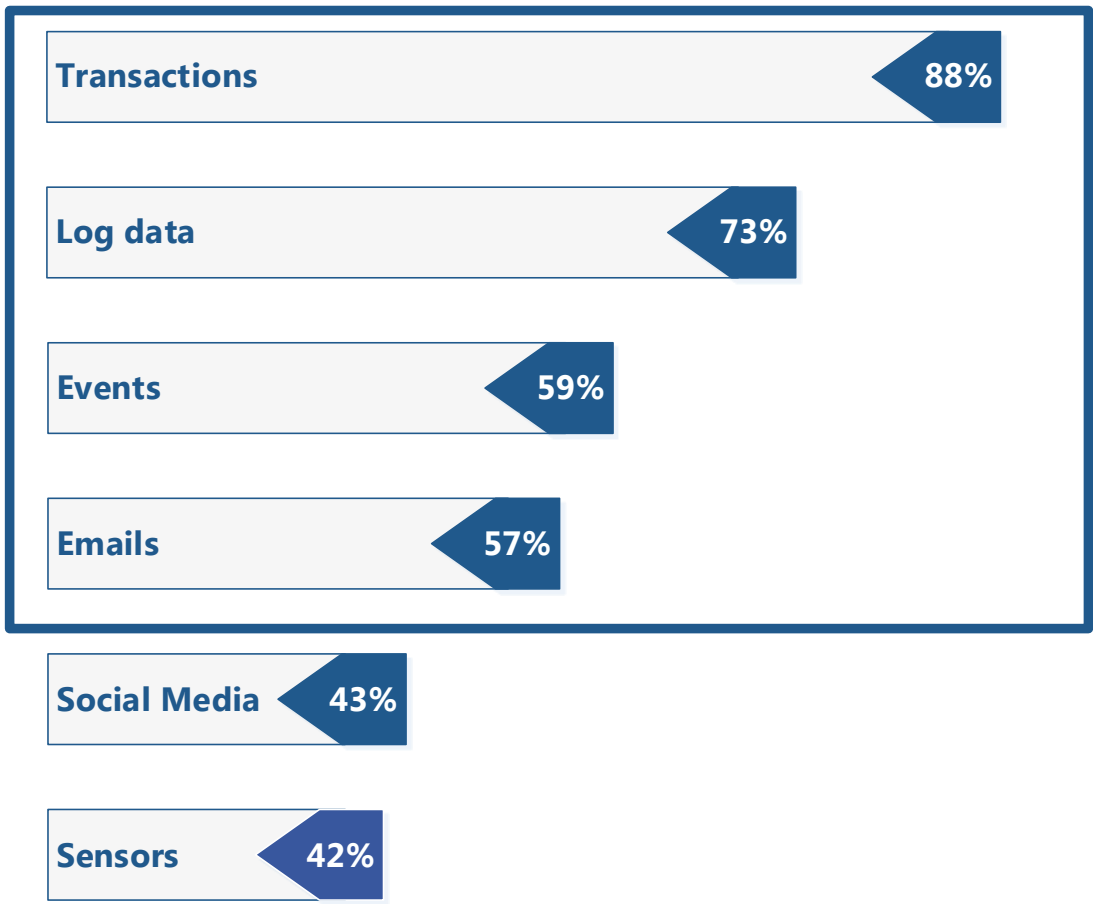
## Dimension: Variety

Big data is any type of data – structured and unstructured data, such as text, sensor data, audio, video, click streams, log files and more, where insights are found when analyzed together:

- Monitor 100's of live video feeds from surveillance cameras to target points of interest
- Exploit the 80 % data growth in images, video and documents to improve customer satisfaction

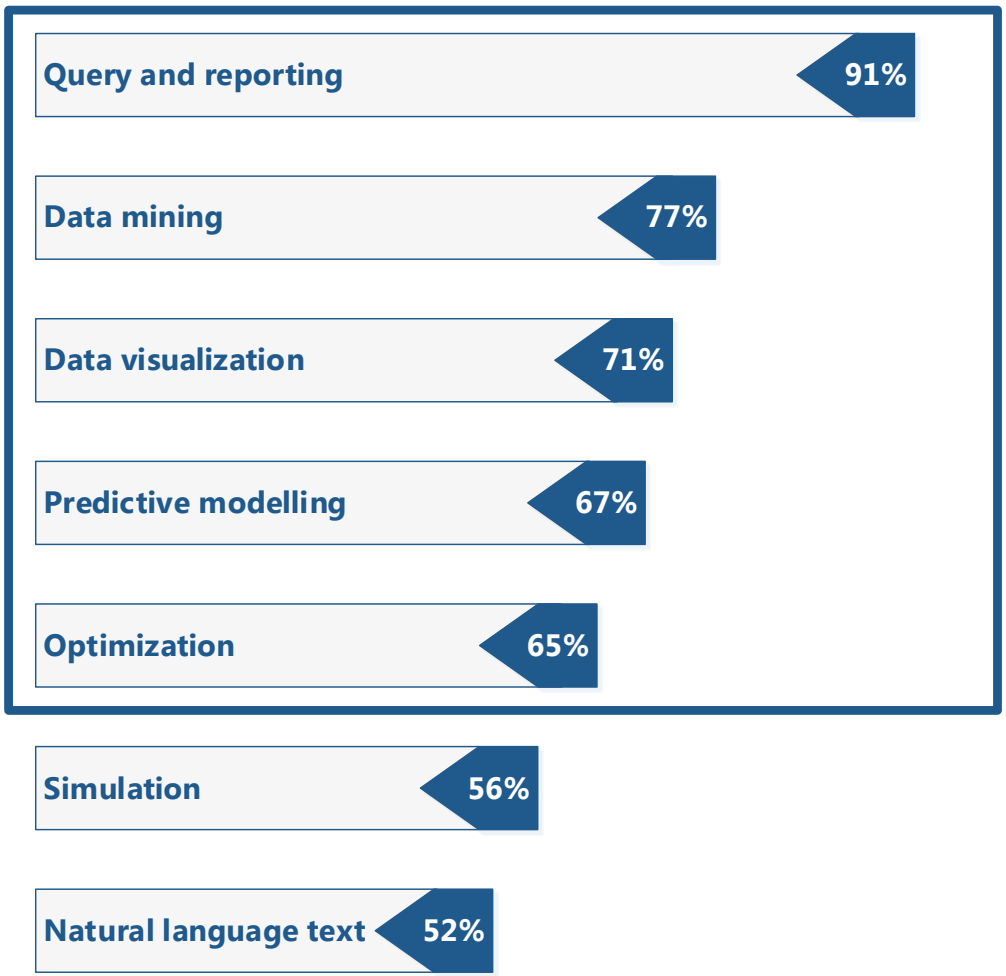
→ IBM (2012). Analytics: The real-world use of big data, Executive Report.

# Data Origin




→ IBM (2012). Analytics: The real-world use of big data, Executive Report.

# Analytics Capabilities




→ IBM (2012). Analytics: The real-world use of big data, Executive Report.

# Successful Examples

| | |
|---|---|
| Healthcare | 20 % decrease in patient mortality by analyzing patient data |
| Telco | 92 % decrease in processing time by analyzing networking and call data |
| Electricity | 99 % improved accuracy in placing power generation resources by analyzing 2.8 petabytes of untouched data |

→ IBM (2012). Analytics: The real-world use of big data, Executive Report.

# Business Intelligence

- Technologies of how information is collected, analyzed and visualized
- Aim at better decision support
- Business Intelligence covers various areas

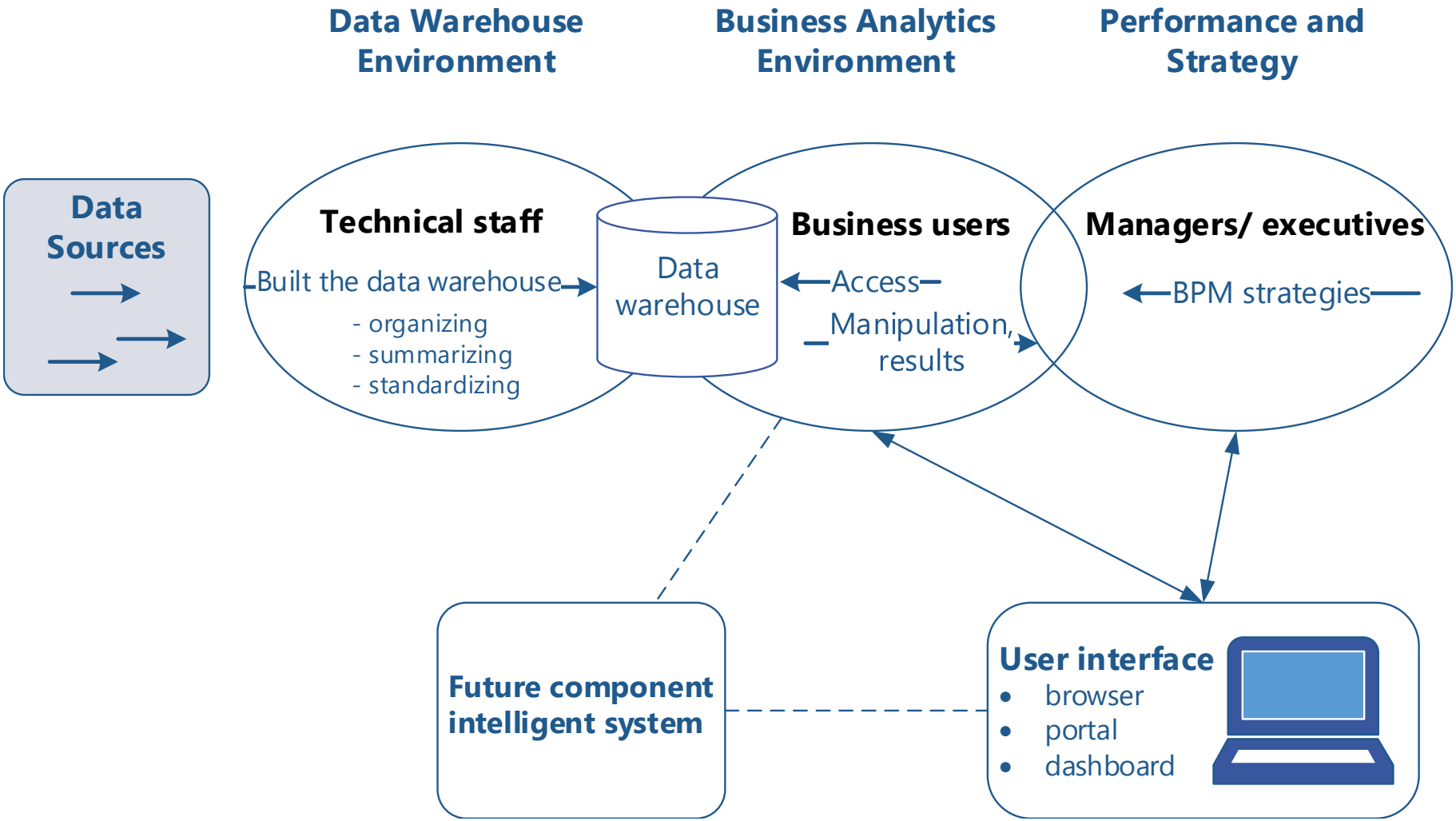
## Covered within Lecture

- Data Analysis
- Forecasting
- Dimension Reduction
- Text Mining

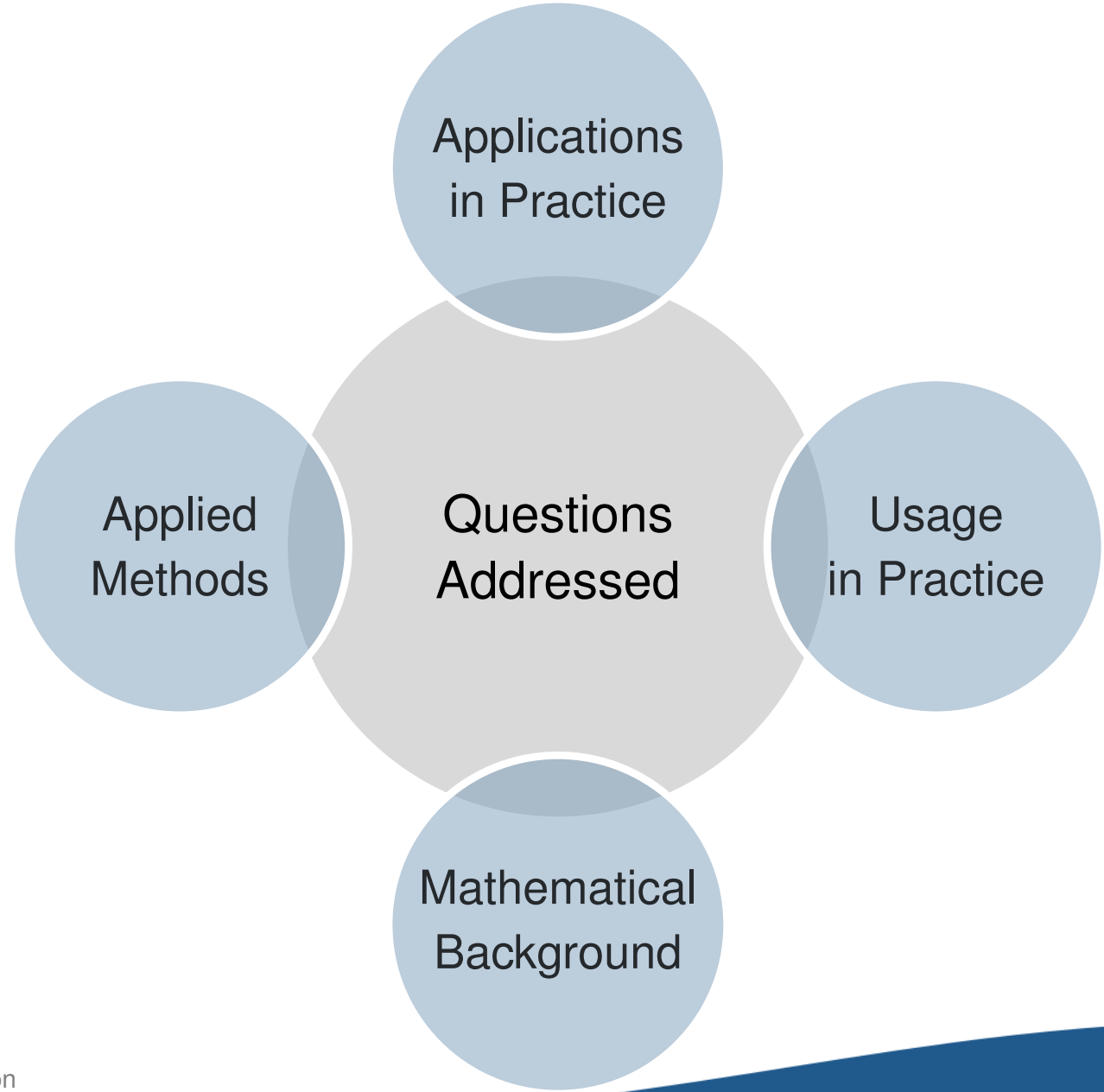
## Beyond Lecture

- Green Computing
- Social Networking
- Cloud Computing
- Multitouch

# High-Level Architecture of Business Intelligence



# Scope of Exercises



# Outline

# Problem Setting

- Can we predict the credit scores of consumers?
- Dataset of 1000 past German applications
  $\rightarrow$ labeled as good (70 %) and bad credit (30 %)
  $\rightarrow$ additional columns, e. g., age, gender, purpose, ...
- Aim: Create a model to predict the probability that applicants have good credit, to reduce the risk to the lender
  $\rightarrow$ Can we beat the baseline accuracy of 70 %?

# Worked Example

```
# load packages
library(caret)
library(e1071)
```

```
# load data
data(GermanCredit)

# train model
set.seed(0)
inTrain <- runif(nrow(GermanCredit)) < 0.2
model <- svm(Class ~ ., data=GermanCredit[inTrain,],
             type="C-classification", kernel="radial")

# predict credit scores
pred <- predict(model, GermanCredit[-inTrain,])
results <- table(pred=pred, true=GermanCredit$Class[-inTrain])

# calculate accuracy
sum(diag(results))/sum(sum(results))


## [1] 0.7558
```

$\rightarrow$ Increase in accuracy from 70 % to 75.58 % yields monetary benefit

# Outlook

1. Introduction to R
2. Data Visualization: Showing relationships graphically
3. Data Analysis: Explaining relationships and patterns statistically
4. Data Mining: Forecasting with machine learning
5. Text Mining

**Question: Which area belongs to Business Intelligence?**

Visit `http://pingo.upb.de` with code 1523